

## Durham Research Online

---

### Deposited in DRO:

03 November 2021

### Version of attached file:

Published Version

### Peer-review status of attached file:

Peer-reviewed

### Citation for published item:

Pöntinen, M. and Granvik, M. and Nucita, A.A. and Conversi, L. and Altieri, B. and Auricchio, N. and Bodendorf, C. and Bonino, D. and Brescia, M. and Capobianco, V. and Carretero, J. and Carry, B. and Castellano, M. and Cledassou, R. and Congedo, G. and Corcione, L. and Cropper, M. and Dusini, S. and Frailis, M. and Franceschi, E. and Fumana, M. and Garilli, B. and Grupp, F. and Hormuth, F. and Israel, H. and Jahnke, K. and Kermiche, S. and Kitching, T. and Kohley, R. and Kubik, B. and Kunz, M. and Laureijs, R. and Lilje, P.B. and Lloro, I. and Maiorano, E. and Marggraf, O. and Massey, R. and Meneghetti, M. and Meylan, G. and Moscardini, L. and Padilla, C. and Paltani, S. and Pasian, F. and Pires, S. and Polenta, G. and Raison, F. and Roncarelli, M. and Rossetti, E. and Saglia, R. and Schneider, P. and Secroun, A. and Serrano, S. and Sirri, G. and Taylor, A.N. and Tereno, I. and Toledo-Moreo, R. and Valenziano, L. and Wang, Y. and Wetzstein, M. and Zoubian, J. (2020) 'Euclid: Identification of asteroid streaks in simulated images using StreakDet software.', *Astronomy Astrophysics*, 644 . A35.

### Further information on publisher's website:

<https://doi.org/10.1051/0004-6361/202037765>

### Publisher's copyright statement:

© ESO 2020

### Additional information:

### Use policy

---

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in DRO
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full DRO policy](#) for further details.

# ***Euclid*: Identification of asteroid streaks in simulated images using StreakDet software<sup>★</sup>**

M. Pöntinen<sup>1</sup>, M. Granvik<sup>1,2</sup>, A. A. Nucita<sup>3,4</sup>, L. Conversi<sup>5,6</sup>, B. Altieri<sup>6</sup>, N. Auricchio<sup>7</sup>, C. Bodendorf<sup>8</sup>, D. Bonino<sup>9</sup>, M. Brescia<sup>10</sup>, V. Capobianco<sup>9</sup>, J. Carretero<sup>11</sup>, B. Carry<sup>12</sup>, M. Castellano<sup>13</sup>, R. Cledassou<sup>14</sup>, G. Congedo<sup>15</sup>, L. Corcione<sup>9</sup>, M. Cropper<sup>16</sup>, S. Dusini<sup>17</sup>, M. Frailis<sup>18</sup>, E. Franceschi<sup>7</sup>, M. Fumana<sup>19</sup>, B. Garilli<sup>19</sup>, F. Grupp<sup>8</sup>, F. Hormuth<sup>20</sup>, H. Israel<sup>21</sup>, K. Jahnke<sup>22</sup>, S. Kermiche<sup>23</sup>, T. Kitching<sup>16</sup>, R. Kohley<sup>6</sup>, B. Kubik<sup>24</sup>, M. Kunz<sup>25</sup>, R. Laureijs<sup>26</sup>, P. B. Lilje<sup>27</sup>, I. Lloro<sup>28,29</sup>, E. Maiorano<sup>7</sup>, O. Marggraf<sup>30</sup>, R. Massey<sup>31</sup>, M. Meneghetti<sup>7</sup>, G. Meylan<sup>32</sup>, L. Moscardini<sup>7,33,34</sup>, C. Padilla<sup>11</sup>, S. Paltani<sup>35</sup>, F. Pasian<sup>18</sup>, S. Pires<sup>36</sup>, G. Polenta<sup>37</sup>, F. Raison<sup>8</sup>, M. Roncarelli<sup>7,33</sup>, E. Rossetti<sup>33</sup>, R. Saglia<sup>8</sup>, P. Schneider<sup>30</sup>, A. Secroun<sup>23</sup>, S. Serrano<sup>28,29</sup>, G. Sirri<sup>38</sup>, A. N. Taylor<sup>15</sup>, I. Tereno<sup>39,40</sup>, R. Toledo-Moreo<sup>41</sup>, L. Valenziano<sup>7,38</sup>, Y. Wang<sup>42</sup>, M. Wetzstein<sup>8</sup>, and J. Zoubian<sup>23</sup>

(Affiliations can be found after the references)

Received 18 February 2020 / Accepted 26 October 2020

## **ABSTRACT**

**Context.** The ESA *Euclid* space telescope could observe up to 150 000 asteroids as a side product of its primary cosmological mission. Asteroids appear as trailed sources, that is streaks, in the images. Owing to the survey area of 15 000 square degrees and the number of sources, automated methods have to be used to find them. *Euclid* is equipped with a visible camera, VIS (VISual imager), and a near-infrared camera, NISP (Near-Infrared Spectrometer and Photometer), with three filters.

**Aims.** We aim to develop a pipeline to detect fast-moving objects in *Euclid* images, with both high completeness and high purity.

**Methods.** We tested the StreakDet software to find asteroids from simulated *Euclid* images. We optimized the parameters of StreakDet to maximize completeness, and developed a post-processing algorithm to improve the purity of the sample of detected sources by removing false-positive detections.

**Results.** StreakDet finds 96.9% of the synthetic asteroid streaks with apparent magnitudes brighter than 23rd magnitude and streak lengths longer than 15 pixels (10 arcsec h<sup>-1</sup>), but this comes at the cost of finding a high number of false positives. The number of false positives can be radically reduced with multi-streak analysis, which utilizes all four dithers obtained by *Euclid*.

**Conclusions.** StreakDet is a good tool for identifying asteroids in *Euclid* images, but there is still room for improvement, in particular, for finding short (less than 13 pixels, corresponding to 8 arcsec h<sup>-1</sup>) and/or faint streaks (fainter than the apparent magnitude of 23).

**Key words.** methods: data analysis – techniques: image processing – minor planets, asteroids: general – space vehicles – surveys – methods: numerical

## **1. Introduction**

European Space Agency's (ESA) upcoming, cosmological *Euclid* mission surveys a large portion of the sky (Laureijs et al. 2011). Even though it purposefully points further than 15° from the ecliptic plane, *Euclid* also detects a large number of Solar System objects (SSO), up to 150 000 (Carry 2018). As the telescope of *Euclid* points at a fixed position of the sky during a measurement, the SSOs move relative to the background sky, and objects moving faster than ~5 arcsec h<sup>-1</sup> (from near-Earth asteroids to Jupiter Trojans) appear as streaks in the data. These fast-moving objects form the majority, approximately two-thirds, of SSOs detected by *Euclid* (see Table 1). Thus, the main problem to be solved is streak detection. The challenges are that the images contain many other sources, such as stars and galaxies, and that most of the linear features in the images are due to cosmic rays rather than asteroids.

The importance of finding the asteroids in *Euclid* images is twofold. First, they offer valuable data for Solar System science. Most of the goals for studying asteroids benefit from understanding their compositions (Gaffey et al. 2002). Observ-

ing the spectral energy distribution of asteroids, in particular in the near-infrared, is essential for their compositional modeling (Reddy et al. 2015). The *Euclid* mission substantially increases the number of asteroids with multi-band photometry that extends to near-infrared. Second, the asteroids need to be identified so that they do not appear as artifacts and interfere with the cosmological data-analysis pipeline, which aims at a very precise measurement of galaxy shapes to determine the amount of weak lensing.

The first step in the analysis of *Euclid* data, in terms of Solar System science, is to find the asteroids in *Euclid* images, because most of the asteroids that are visible in the images are previously undiscovered objects (see Table 1). Lieu et al. (2019) carried out tests using deep convolutional neural networks (CNNs) to find the streaks in *Euclid* images. A related method has been developed by Duev et al. (2019) to use an ensemble of CNNs to remove false-positive streaks in Zwicky Transient Facility data. The CNN approach is auspicious; however, currently it has only been used to classify small (up to a few hundred pixels wide) sub-images containing streak candidates, but not to extract streak coordinates from large images. Furthermore, even though the simulated *Euclid* images imitate the future real data as closely as possible, the real images might look different. Because of the gap

<sup>★</sup> This paper is published on behalf of the Euclid Consortium.

between synthetic and real image distributions, a CNN that is only trained with simulated training data probably does not work optimally out of the box for the real images (Shrivastava et al. 2017). If machine learning is used, it is advisable to have other algorithms for finding the asteroids, if for nothing else, at least for gathering non-synthetic training examples for the neural networks. There is a parallel effort to develop a non-machine-learning pipeline from simulated *Euclid* images by Nucita et al. (in prep.), which is focused on finding streaks that are shorter than 15 pixels ( $10 \text{ arcsec h}^{-1}$ ). Other astronomical streak detection methods are, for example, findStreaks (Waszczak et al. 2017) and a method based on fast Radon transform (Nir et al. 2018).

The method used in this paper to tackle the problem of finding the asteroids, especially the faster-moving ones, is a software called StreakDet, which was developed by Virtanen et al. (2016) to detect streaks caused by space debris in images obtained from an Earth-orbiting platform. We tested what proportion of simulated SSOs can be detected from the synthetic images with StreakDet, what the detectable apparent magnitude and sky motion ranges are, and how high completeness and purity can be achieved. We did this by optimizing the parameters of StreakDet, developing a post-processing step to decrease the number of false positives, and programming a test and analysis software to give statistics on the results. The simulated data were generated with ELViS, a Python program developed within the Euclid Consortium.

In what follows, we first describe the *Euclid* mission and its predicted main contributions to Solar System science. Then, we describe the simulation of *Euclid* VIS images, the methods for extracting streaks from images, and the tools for analyzing the results. We then present and discuss our results obtained with the StreakDet software and our additional post-processing step. Finally, we provide our conclusions.

## 2. Euclid

### 2.1. Euclid mission

*Euclid* is a medium-class mission in the Cosmic Vision program of ESA. It was selected in October 2011 and is planned to launch in 2022. The nominal science program duration is six years, a time during which *Euclid* stays in a halo orbit around the second Lagrange point of the Sun-Earth system. The mission is named after the mathematician of ancient Greece, Euclid of Alexandria.

*Euclid* is a cosmology mission, and it surveys a large portion of the sky,  $15\,000 \text{ deg}^2$  (Laureijs et al. 2011). *Euclid* explores the nature of dark energy, dark matter, and gravity using gravitational lensing effects on galaxies (weak lensing) and the properties of galaxy clustering (baryonic acoustic oscillations and redshift space distortion) (Amendola et al. 2018).

*Euclid* has a telescope with an aperture of 1.2 m and a focal length of 24.5 m (Venancio et al. 2014). The mission operates in the wavelength range from 550 nm (green) to 2000 nm (near-infrared). The measuring instruments are VIS (VISual imager) and NISP (Near-Infrared Spectrometer and Photometer), sharing the same  $0.53 \text{ deg}^2$  field of view. VIS is a 600-megapixel visible imager, operating between the wavelengths of 550 and 900 nm (Cropper et al. 2018). NISP consists of a near-infrared three-filter photometer (NISP-P) and a slitless spectrograph (NISP-S) with a resolving power of 380 (Maciaszek et al. 2016). The pixel size for VIS is 0.1 arcseconds per pixel, and for NISP 0.3 arcseconds per pixel. *Euclid* operates in a step-and-stare mode: VIS and NISP-S observe an area of the sky simultaneously with an

exposure time of 565 seconds, after which NISP-P executes three shorter exposures with *Y*, *J*, and *H* filters for 121 s, 116 s, and 81 s, respectively (Racca et al. 2016). For a given field, the aforementioned exposures are repeated four times, with small changes in the telescope pointing direction in between them, so that the total observing time for each field is approximately 70 min. The current baseline now also includes a fifth VIS image, taken simultaneously with the NISP/J photometry exposure, during the third jitter.

Since the data created by *Euclid* consist of hundreds of thousands of images, with an accumulated data volume of several petabytes, automated algorithms are necessary to analyze the data. Approximately 10 billion galaxies appear in the data, of which over 1 billion are used for weak-lensing, and several tens of millions of galaxy redshifts are measured (Laureijs et al. 2011).

### 2.2. Solar System objects in Euclid images

Although *Euclid*'s main science goals are cosmological, it also observes up to 150 000 Solar System objects (Carry 2018). Most of these are asteroids, and the ones faster than  $\sim 5 \text{ arcsec h}^{-1}$  (orbits approximately up to Jupiter's orbit) appear in the *Euclid* images as streaks. The detection limit with VIS for  $10\sigma$  on  $0.43 \text{ arcsec}$  extended sources is around  $m_{\text{AB}} = 24.9$  (Cropper et al. 2018), and with NISP *Y*, *J*, and *H* filters for  $5\sigma$  point sources, it is around  $m_{\text{AB}} = 24$ . The NISP slitless spectra is obtained by two grisms, blue ( $0.92$  to  $1.25 \mu\text{m}$ ), which is used only in the deep survey, and red ( $1.25$  to  $1.85 \mu\text{m}$ ), which is used both in wide and deep surveys. The grisms have a continuum sensitivity of around  $m_{\text{AB}} = 21$ .

Detailed estimates of the observed SSOs are shown in Table 1. In addition to the object classes shown in Table 1, *Euclid* has the potential to observe interstellar objects (ISO) as well. The *Vera C. Rubin* Observatory (VRO; formerly known as the Large Synoptic Survey Telescope) is estimated to find on the order of ten, possibly even more, ISOs during its ten-year survey (Cook et al. 2016). *Euclid* covers a slightly smaller area of the sky than VRO, and the survey is shorter, so as a crude estimate, the number of detected ISOs is expected to be a bit smaller than VRO. All in all, *Euclid* can make a viable contribution to the detection of ISOs, especially in the northern sky, because VRO covers the southern sky only.

*Euclid* avoids the Galactic and ecliptic planes by observing the sky that has Galactic latitudes greater than  $30^\circ$ , and ecliptic latitudes higher than  $15^\circ$ , except for calibration fields, which can be closer to the ecliptic plane. For this reason, the asteroids detected by *Euclid* are primarily objects in high-inclination orbits, which are currently underrepresented in the census of asteroids (Mahlke et al. 2018). Nevertheless, the calibration field data close to the ecliptic plane contain the highest number of asteroids per field of view, up to a few thousand. There are approximately 60% more known asteroids for every 3 degrees closer to the ecliptic. There are also some asteroids in the deep fields, but not many because they are far away from the ecliptic.

As *Euclid* measures multi-band photometry of galaxies, it also substantially increases the number of asteroids with multi-band photometry in the near-infrared region. In addition to the compositional information provided by multi-band photometry, *Euclid* data can, in many cases, also be used to constrain several other properties of the asteroids, such as the rotation period, spin-axis orientation, and shape, as well as detect binary asteroids (Carry 2018). Due to the relatively short observational time span per asteroid, accurate orbit solutions for the discoveries

**Table 1.** Estimates of *Euclid* survey parameters for different kinds of Solar System objects.

Population	Known objects	<i>Euclid</i> discoveries	<i>Euclid</i> observations	$H_V$ magnitude limits			arcsec h <sup>-1</sup>	VIS Pixels
NEA	24 086	$1.4^{+1.0}_{-0.5} \times 10^4$	$1.5^{+1.0}_{-0.6} \times 10^4$	22.75	23.75	26.50	$43.3^{+36.5}_{-19.9}$	$67.9^{+57.2}_{-31.2}$
MC	16 662	$1.0^{+1.7}_{-0.8} \times 10^4$	$1.2^{+1.7}_{-0.8} \times 10^4$	21.00	21.25	22.75	$41.3^{+22.6}_{-14.9}$	$64.8^{+35.4}_{-23.3}$
MB	937 295	$8.2^{+2.5}_{-2.2} \times 10^4$	$9.7^{+2.5}_{-2.2} \times 10^4$	19.50	20.00	21.25	$32.5^{+7.9}_{-5.5}$	$51.0^{+12.4}_{-8.6}$
Trojan	8846	$7.1^{+9.3}_{-4.9} \times 10^3$	$7.5^{+9.5}_{-5.0} \times 10^3$	17.00	17.25	18.25	$13.3^{+1.4}_{-1.1}$	$20.9^{+2.2}_{-1.7}$
Centaur	951	$2.2^{+2.1}_{-1.4} \times 10^3$	$2.2^{+2.1}_{-1.4} \times 10^3$	14.75	15.50	18.25	$4.0^{+2.9}_{-1.5}$	$6.2^{+4.5}_{-2.3}$
KBO	2553	$5.3^{+1.6}_{-1.3} \times 10^3$	$5.5^{+1.6}_{-1.3} \times 10^3$	8.25	8.75	10.00	$0.6^{+0.3}_{-0.1}$	$1.0^{+0.5}_{-0.2}$
Comet	4070	$22^{+4}_{-4}$	$38^{+5}_{-4}$	18.25	19.00	22.00	$4.4^{+6.2}_{-1.8}$	$6.9^{+9.7}_{-2.8}$

**Notes.** NEA stands for near-Earth asteroid, MC for Mars-crossing, MB for main-belt, Trojan for Jovian Trojan, and KBO for Kuiper-belt object. Column 4 (*Euclid* observations) shows the estimated total number of Solar System objects in the data, while Col. 3 (*Euclid* discoveries) shows the estimated number of previously unknown observed objects. Columns 5, 6, and 7 ( $H_V$  magnitude limits) show the largest absolute magnitudes  $H_V$ , for which the observation probability is 100%, 50%, and 1%, respectively. The values were determined from simulations, and they show the likelihood that an object of a given absolute magnitude appears in *Euclid* images so that the apparent magnitude is enough to observe the object. Column 8 (arcsec h<sup>-1</sup>) shows the median sky motion of the objects, with 25th and 75th percentiles in the subscripts and superscripts, and Col. 9 (VIS pixels) shows the corresponding streak lengths on the VIS CCD. The data are from [Carr \(2018\)](#) except for Col. 2, for which the numbers of currently known objects were updated by the Minor Planet Center, as of October 17, 2020.

cannot be obtained by using *Euclid* data only, but rough orbits and especially inclination distributions can be estimated (cf. [Muinonen et al. 2006](#)). More accurate orbits can be obtained if additional astrometry is or becomes available. For example, the upcoming VRO determines the orbits for a large number of asteroids, and these orbits can be cross-correlated with *Euclid* detections (cf. [Snodgrass et al. 2018](#)).

### 3. Simulated *Euclid* images and preprocessing

The simulated data are generated with ELViS, a Python program developed within the Euclid Consortium, to mimic the actual *Euclid* VIS data as closely as possible. Here, we focus on the simulated VIS data, as it is the instrument with the higher sensitivity, longer exposure time, and smaller pixels, thus being better suited for the StreakDet software. Simulated VIS images include astronomical sources such as galaxies, stars, and Solar System objects as well as image artifacts such as cosmic rays (CRs), ghosts, charge transfer inefficiency (CTI), and bleeding. The images also contain Poisson noise, readout noise, and bias. See the top left image of Fig. 1 for an example of the data.

The field we simulated is centered at Galactic latitude 85.0° and longitude -82.8°. The number of galaxies per CCD range from 469 to 822, with a mean of 653. For stars, the range is from 78 to 126, and the mean is 100 stars per CCD. The simulation galaxy input catalog was created by the *Euclid* cosmology science working group based on the MICE2 simulation ([Crocce et al. 2015](#)), for the usage of *Euclid* “Science Challenge 3”. Due to undocumented reasons, probably related to processing volume, the galaxy count is approximately half of what is expected for the real sky, and galaxies fainter than 24.5 magnitudes are left out.

The asteroids in our data set were simulated to come from a uniform random distribution between apparent magnitudes 20 and 26, apparent velocities between 10 and 80 arcsec h<sup>-1</sup>, and angles between 0 and 360 degrees (clockwise from east), except for one image set that focused more on shorter streaks, with velocities ranging from 1 to 20 arcsec h<sup>-1</sup>. On average, there are 25 simulated asteroid streaks in each CCD. The simulated SSO population is in no way realistic, but this flat synthetic population is more suited for analyzing the performance metrics of StreakDet and our pipeline.

Due to a reason that has yet to be identified within the ELViS software, there is a small systematic, visible offset (up to a few pixels) between the intended coordinates of the simulated asteroids and the final positions of the simulated asteroid streaks. The offset is a linear function of right ascension and declination. Therefore, we applied a linear correction to the ground-truth catalogs to correct for the systematic offset.

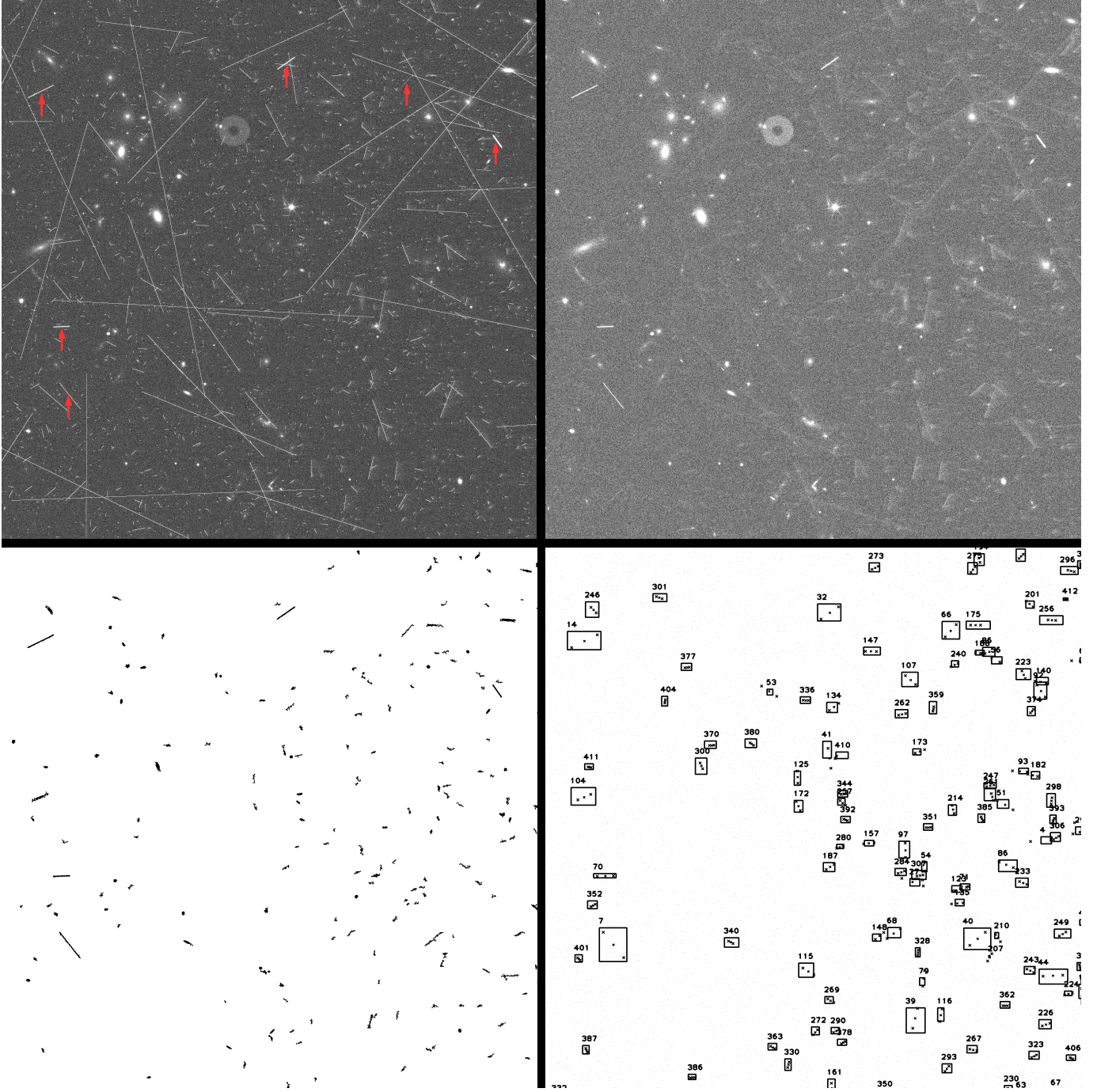
The simulated data are provided as multi-extension FITS files, and each file contains a header and four 2048 × 2066-pixel quadrants of a single 4096 × 4132-pixel CCD. One simulated exposure consists of 36 such CCDs, forming a 6 × 6 grid. Between exposures, there is a dither movement, and one data set consists of four dithered exposures, totaling 144 CCDs. The simulated data did not include the fifth shorter VIS exposure. We run tests on ten such sets for faster moving objects, and one set for slower objects, totaling 11 sets, adding up to a total of 1584 CCDs.

Before feeding the images to StreakDet, we tiled the 2 k × 2 k-pixel image quadrants of multi-extension FITS files into single-extension 4 k × 4 k-pixel images. We also removed CRs with Astro-SCRAPPY ([McCully et al. 2018](#); [Van Dokkum 2001](#)). Astro-SCRAPPY removed essentially all bright pixels caused by cosmic rays. Some residuals were visible in the images afterward, which was mainly caused by the CTI effect (see top right image in Fig. 1). Astro-SCRAPPY does not remove asteroid streaks due to their different PSF shape compared to cosmic-ray streaks.

The data reduction for actual *Euclid* images is handled by *Euclid* OU-VIS (Organizational Unit VIS) with more advanced pipelines than were used in this work. The OU-VIS reduction pipeline includes bias subtraction, flat fielding, CR removal, and CTI corrections, as well as astrometric and photometric calibrations with the *Gaia* ([Gaia Collaboration 2016](#)) catalog ([Dubath et al. 2016](#)).

Most of the asteroids appear in three or four dithers of the simulated data. Exceptions include asteroids that are close to the edge of an image in first dither, which subsequently move out of the imaging field due to their apparent velocity or the dither movement of the imaging field. Similarly, asteroids that are outside the imaging area in the earlier dithers can move into the images in later dithers. Thus, some asteroids appear in as little as one dither. For our simulated data, in which the apparent





**Fig. 1.** Streak detection steps. The *top left* image shows a quadrant of a raw CCD file. The size of the image is  $2048 \times 2066$  pixels, corresponding to  $204.8 \times 206.6$  arcseconds. Asteroids are marked with red arrows, and the remaining streaks are cosmic rays. The *top right* image is the same file after having cosmic rays removed. This image is then fed to *StreakDet*, which results in a segmented image, shown in the *bottom left*. The *bottom right* image shows the final streaks found by *StreakDet*. As can be seen, there are many false-positive detections. The brightest pixels in an image are often caused by cosmic rays, and after they are removed, the scaling of the image changes. For this reason, after cosmic ray removal, the image appears brighter than before the cosmic rays were removed.

velocities, magnitudes, and streak angles come from a uniform random distribution, the visibility of asteroids was as follows: 37% for four streaks, 45% for three streaks, 9% for two streaks, and 9% for one streak. For a realistic asteroid population distribution, the visibility statistics are likely to differ. Nevertheless, for most objects, it is not necessary to find all the streaks in all dithers caused by an object, but as little as one out of four can count as a find, or as little as two when utilizing the post-processing step to link streaks from several dithers.

## 4. Methods

### 4.1. *StreakDet*

*StreakDet* (Virtanen et al. 2016) is an ESA-funded software developed to detect and analyze object trails in imaging data. *StreakDet* has been developed mainly to analyze observations of space debris either from Earth-based or space-based platforms, and it can detect long, faint, and also curved streaks. Its focus is on being able to detect streaks from single images as

opposed to finding consecutive streaks from stacked images, a task for which it is not well optimized. The initial tests run (on non-*Euclid* images) by the StreakDet developers gave detection sensitivities of about 90% for bright streaks (per pixel signal-to-noise ratios of  $>1$ ), and about 50% for dimmer streaks ( $S/N = 0.5$ ).

The StreakDet pipeline consists of three main phases: segmentation, classification, and lastly astrometric and photometric reduction. The following descriptions of the algorithms are summarized from Virtanen et al. (2016). In the following description, black and white (BW) refers to a binary image with a color depth of 1 bit, where every pixel has either a value of 0 or 1. Grayscale (GS) refers to the original image with a higher bit depth.

The segmentation step converts the analyzed image into BW to make the following steps computationally less demanding. The BW image is created with the aid of two GS mean filters. The first filter uses a smaller area, for example,  $3 \times 3$  pixels, for which it calculates a mean pixel value. The second filter uses a larger area, for example,  $21 \times 21$  pixels, and calculates the mean pixel value for that area. The BW image is created by subtracting the differences of the means calculated by the GS mean filters. The idea behind the mean-differences is to detect groups of pixels whose value differs from the background. Because the mean-difference calculation is locally executed, it is typically not biased by global background gradients, which reduces the need for preprocessing the image before feeding it to StreakDet. Before segmentation, the density of bright pixels in the image is calculated or manually set in order to determine the proportion of white pixels to black pixels in the segmented image.

After the segmentation has turned the image to black and white, filtering processes are applied, which aim to remove non-streak-like features from the pipeline. An adapted version of binary erosion is used, which gets rid of isolated active pixels and keeps pixels which are part of a larger structure, such as a streak. Then a reconstruction filter is used to strengthen the remaining features. After the previous steps have removed most of the noise and small stars, larger stars are removed by multiple-window pixel-density evaluation, which removes pixel groups whose number of active pixels does not grow linearly when the window size is increased. Finally, the remaining features and their properties are indexed into a list with a connected component labeling (CCL) algorithm.

After segmentation, the CCL features are given as input into the classification phase. This consists of the following three steps: first, the characterization of BW CCL features; second, the characterization of BW CCL features that correspond to streaks; third, the characterization of original GS features that correspond to streaks.

Each of the steps uses classification processes to find the streaks, and filtering processes to eliminate the non-streaks. During each step, eigenvalue analysis is used to compute feature parameters such as width, orientation angle, aspect ratio, curvature, and porosity (referring to the compactness of the feature). Both linking and unlinking are done during the BW steps. Linking implies connecting found features that are likely to be parts of the same longer streak. In contrast, unlinking implies dividing a large found feature into smaller ones, if it seems likely that the sub-features are actually separate streaks. During the GS parametrization, point-spread-function (PSF) fitting with a moving 2D Gaussian approach, utilizing the nonlinear Levenberg-Marquardt least-squares method, is used to refine the streak parameters. The GS parametrization is done by starting with the BW features and then finding the parameters of the cor-

responding GS streak with the aforementioned algorithms. The GS fitting also extends outside the BW bounding box, in case only a part of the real streak was found during the BW process. Finally, the GS features are classified and filtered according to their PSF width and curvature. An example of the results after segmentation and after the whole pipeline can be seen in the bottom left and bottom right images of Fig. 1, respectively. The final, optional phase of the StreakDet pipeline is astrometric and photometric reduction.

Because StreakDet uses only one CPU most of the time,  $N$  StreakDet processes can be run in parallel with a computer with  $N$  CPU cores, which makes it possible to run StreakDet on 36 FITS CCDs in only approximately  $\max(1, 36N^{-1})$  times the time it takes to analyze one FITS CCD. With our settings, running StreakDet on a  $4k \times 4k$  image using a single CPU core took approximately one minute. It takes *Euclid* 70 min to produce the 180 VIS CCDs of one pointing, so, theoretically, three CPU cores would be enough to run StreakDet with VIS data so that the program would not fall behind the image delivery pace of *Euclid*. In practice, more CPUs are needed to rerun StreakDet with different parameters or on differently preprocessed VIS images or, alternatively, depending on the schedule at which VIS data become available.

#### 4.2. Analysis program

In order to compare the results given by StreakDet to the simulation inputs (ground truth), we developed a test and analysis program in Python. The analysis program was used to process all the ground-truth and StreakDet data, compute statistics on the results, and show plots of hits, misses, false positives, and errors.

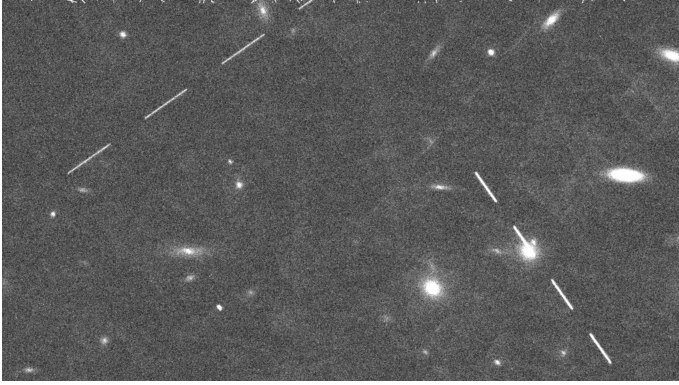
We tested StreakDet by running it on single  $4k \times 4k$ -pixel FITS CCDs, and only afterward did we combine and stack all the data by the separate Python analysis program. The reason for this is that StreakDet itself does not appear to work well with stacked images, that is, the detection percentage was lower for stacked images than for individual images. The reason for which seems to be due to combined noise, causing the signal-to-noise ratio ( $S/N$ ) of the streaks to decrease. The  $4k \times 4k$  images are run one by one with StreakDet, and then the images, StreakDet results, and ground-truth data are tiled and stacked afterward with the analysis program.

To improve the purity of the sample (i.e., to decrease the number of false-positive detections), we developed a post-processing step called multi-streak analysis and implemented it in Python. As the data have images from four dithers, the asteroids appear as multiple line segments along the same line in the stacked image (see Fig. 2). Before feeding the data to the multi-streak pipeline, so-called static streaks are removed. In other words, streaks that are in the same coordinates in multiple dithers cannot be asteroids, but they are galaxies or other static objects instead and thus can be removed from the analysis. With our optimized StreakDet parameters, approximately one quarter of all streaks found by StreakDet were static.

The multi-streak pipeline is described below.

1. Run StreakDet separately on all FITS CCDs for all four dithers.
2. Tile and stack ground-truth data with the Python analysis program.
3. Tile and stack StreakDet results with the Python analysis program.
4. Search for ground-truth multi-streaks that have 2–4 streaks along the same line, so that the streaks are different dithers





**Fig. 2.** Example of asteroids appearing as multi-streaks, when all four dithers are stacked into a single image. The image shows two dashed lines, each of which is caused by an asteroid. The width and height of the image are approximately 1000 and 600 pixels, corresponding to 100 and 60 arcsec, respectively.

in logical order, that is, assuming that the object moves fairly linearly in one direction and does not suddenly move backward.

5. Search for *StreakDet* multi-streaks that have 2–4 streaks along the same line (again, all streak in different dithers should be in logical order).
6. Check that the lengths, angles, fluxes, and PSF sigma (standard deviation from PSF fitting) values of the individual streaks in the multi-streaks are within given ranges of each other.
7. Analyze which *StreakDet* multi-streaks match ground-truth multi-streaks and which do not. To be classified as a match, at least two single streaks in a *StreakDet* multi-streak have to match to those of the ground-truth multi-streak.

Without the ground-truth data available, that is, when processing the real *Euclid* data, only steps 1, 3, 5, and 6 are carried out.

Asteroids that move from one CCD to another between dithers are taken into account in the multi-streak analysis. For example, if an asteroid appears in one CCD in the first dither, and moves to the neighboring CCD in the subsequent dithers, the multi-streak analysis can link all of the separate streaks together.

## 5. Results

### 5.1. Single-streak analysis

For *StreakDet*, the fraction of true positives and especially false positives changes substantially depending on the settings used. After testing the parameters available for the *StreakDet* pipeline, we identified a few parameters that increase the number of true positives. In general, the segmentation settings are the most important part of optimizing *StreakDet*, because if a streak is not already found in the segmentation phase, it becomes impossible to find it during the subsequent processing stages. After optimizing the segmentation phase, parameters related to the later stages of the pipeline were optimized as well in order to get the final detection percentage as close to the segmentation detection percentage as possible. After testing and optimization, a combination of settings appearing to maximize the number of true positives was chosen and used to process all simulated *Euclid* images.

As expected, there is a trade-off between maximizing the number of true positives and minimizing the number of false

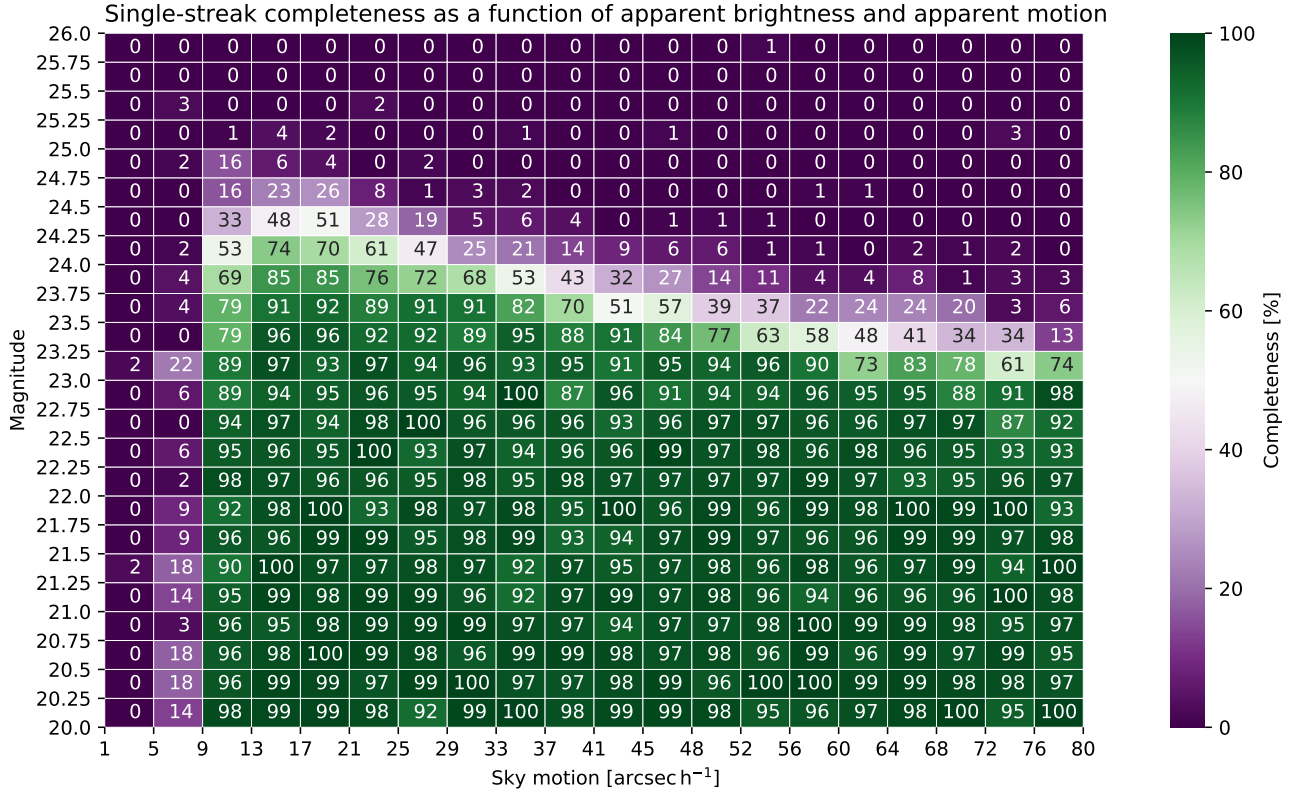
positives. Our primary goal was to maximize the number of true positives during *StreakDet* processing because the number of false positives can be radically reduced with the multi-streak analysis that is carried out as a post-processing step. The optimal parameters seem to vary slightly from image to image, and they may depend on the exact image characteristics, such as background brightness and stellar density. However, *StreakDet* takes the number of bright stars and pixels in the image into account before starting the segmentation, so the variation in results between images is small, at least with the simulated data. It is expected that *StreakDet* can achieve fairly similar results on different parts of the sky, using the same parameters. When aiming for the highest completeness on atypical fields, such as the calibration fields, tuning the parameters might prove useful.

*StreakDet* can differentiate between streaks caused by asteroids and CRs due to their difference in PSF shape. However, when tested on data containing CRs, the total detection percentage achieved by *StreakDet* was approximately 10 percentage points lower compared to data that had CRs removed. Due to this effect, we opted to remove CRs from the data before analyzing it with *StreakDet*. Furthermore, the CRs are removed anyway during the processing of the real *Euclid* data by OU-VIS.

Figure 3 shows the *StreakDet* detection percentage (completeness) for streaks of different magnitudes and lengths as a heat map. The completeness is defined here as the fraction of streaks found among the streaks that are at least partially visible in the CCDs. In other words, streaks that are completely outside or between the CCDs are ignored. All in all, *StreakDet* managed to find 96.9% of individual streaks brighter than 23 magnitudes and longer than 15 pixels (16283 out of 16807). For fainter streaks, the detection percentage starts to fall, reaching essentially zero at magnitude 25. This is to be expected, as the limiting magnitude in the simulated data is 24.5 at an S/N of 10 for an extended source. This approximately corresponds to a magnitude of 26.5, using an S/N of 3 to 5 for a point source. Since the light is spread onto, for example, 10 pixels, a magnitude 24 SSO has approximately the same per-pixel counts as a magnitude 26.5 star, and it becomes difficult to discern from the background. For short streaks, the detection percentage decreases rapidly for streaks shorter than 13 pixels (8 arcsec  $h^{-1}$ ). For fast-moving faint asteroids, the streaks appear fainter than for slow-moving asteroids of the same magnitude because, in a longer streak, the integrated flux is spread over a larger number of pixels. For this reason, shorter streaks caused by faint asteroids are found more easily than longer streaks caused by faint asteroids. Since the simulated galaxy density is only half of what is expected of the real sky and the numerous faint galaxies are missing from the images, the detection percentage for faint streaks could be somewhat different for the actual data. Nevertheless, as *StreakDet* does not detect streaks fainter than magnitude 24.5, the difference in results should not be very significant.

Each SSO forms from one to four streaks, because there are four dithers. Therefore, it is not necessary to detect each streak that formed by an SSO in order to be able to detect that particular object. When calculating the number of simulated SSOs for which *StreakDet* found at least one streak out of up to four streaks, the detection percentage is 99.1% for SSOs brighter than 23 magnitudes and longer than 15 pixels (10 arcsec  $h^{-1}$ ). Again, the detection percentage starts to fall after magnitude 23, but it is 10–20 percentage points higher for SSOs than for individual streaks between magnitudes 23 and 24.5.

According to our tests, *StreakDet*'s detection ability declines rapidly for streaks shorter than approximately 13 pixels



**Fig. 3.** StreakDet detection percentage as functions of apparent magnitude and apparent motion. The values on both axes show the limits of the bins. For example, the bin between sky motions of 33 and 37 arcsec h<sup>-1</sup> and magnitudes of 21.5 and 21.75 shows that the StreakDet detection percentage (completeness) is 99% for streaks created by all simulated objects between those values. On average, there are 83 ground-truth streaks per bin.

(8 arcsec h<sup>-1</sup>). Therefore, for the ten main image sets used for the final results, the lengths of asteroid streaks were chosen to range from 15 to 125 pixels (from 10 to 80 arcsec h<sup>-1</sup>). Also, with certain settings, StreakDet does not appear to find streaks shorter than 15 pixels (10 arcsec h<sup>-1</sup>) and brighter than 20.5 magnitudes at all. This is due to StreakDet being optimized for relatively faint streaks by default. When the pixel brightness crosses a certain threshold, it causes bright streaks to already be filtered out in the segmentation phase. This can be avoided by adjusting parameters in the source code of StreakDet's segmentation phase. Optimizing for short and bright streaks causes the completeness to fall slightly, approximately 0.5 percentage points for all streaks fainter than 21 magnitudes. This is notable because most of the found asteroids are faint, and a small change in the detection percentage close to the limiting magnitude radically changes the number of observed objects. It seems that one set of parameters is good for finding the brightest streaks, and another is better for finding fainter ones. For the real *Euclid* images, it might make sense to run StreakDet twice for all data, with two different parameter sets. For the results in this work, we used the setting optimized for brighter streaks, because it offered a slightly higher total completeness with our uniform data set.

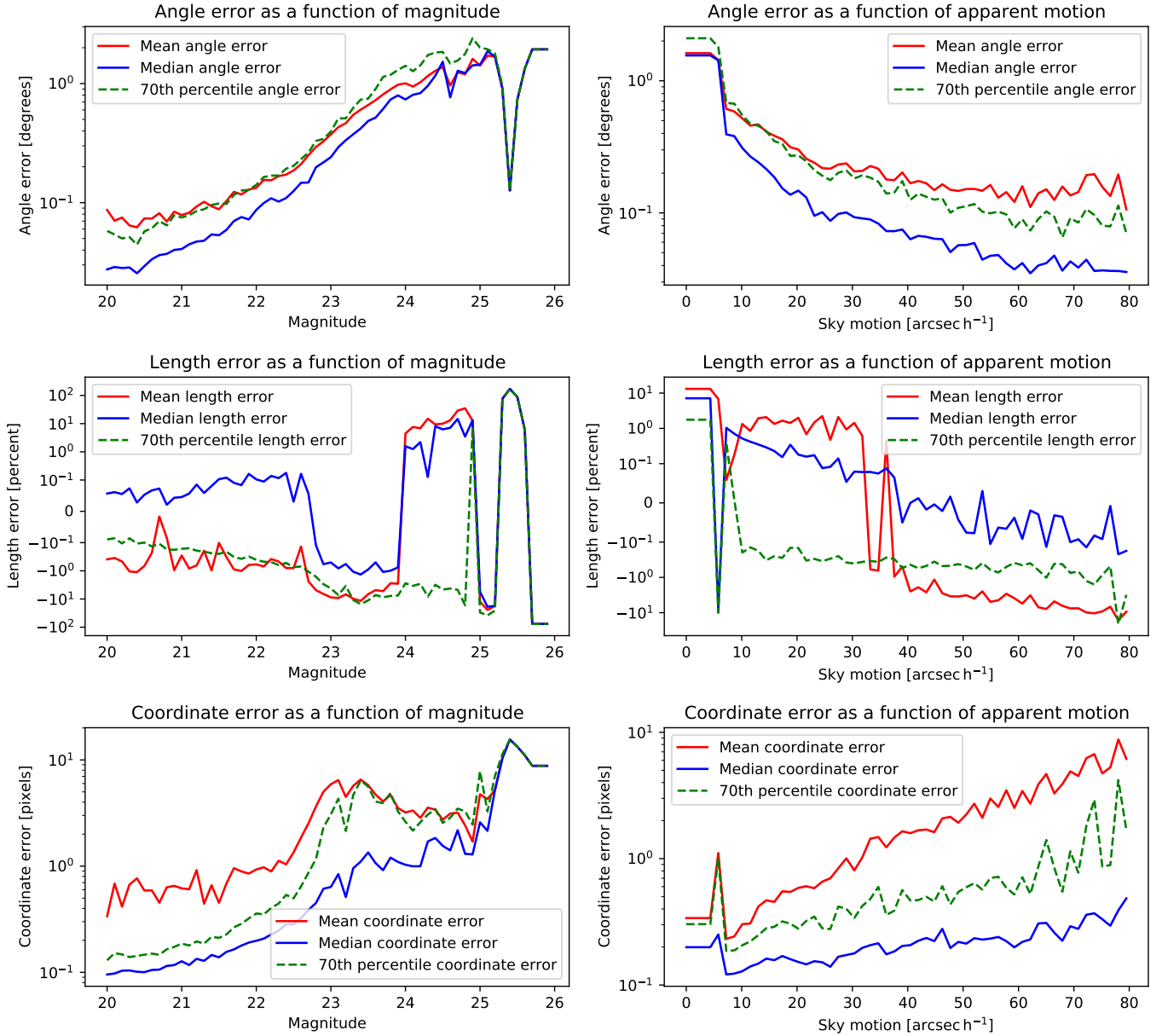
In addition to the ten main image sets focused on streaks longer than 15 pixels (10 arcsec h<sup>-1</sup>), we ran supplementary tests with a simulated data set containing shorter streaks, from approximately 1 pixel to 30 pixels (1–20 arcsec h<sup>-1</sup>). We did this to get a better idea of the streak length range at which StreakDet can operate. According to those tests, StreakDet's detection percentage starts to drop sharply for streaks shorter

than 13 pixels (9 arcsec h<sup>-1</sup>), reaching essentially zero below 9 pixels (6 arcsec h<sup>-1</sup>).

Comparing the results in Fig. 3 to the populations shown in Table 1, in terms of sky motion, StreakDet can find a vast majority of bright enough near-Earth asteroids, Mars-crossing asteroids, main-belt asteroids, and Jovian Trojans (for a visualization of the sky motion ranges of different populations, see Fig. 5 in Carry 2018). In addition, individual Centaurs, Kuiper-belt objects, and comets can form long enough streaks for StreakDet to detect. For the populations shown in Table 1, the apparent magnitude limit is 24.5, which StreakDet can reach for the shortest found streaks (Trojans) with a completeness of approximately 50%. For faster moving objects, such as main-belt asteroids, the limiting magnitude of StreakDet is 24 or less. This suggests that StreakDet cannot detect all of the up to 150 000 objects appearing in the images, but getting precise completeness and purity values for realistically simulated SSO populations requires new simulations and remains for future work.

The brightness of each simulated asteroid stayed constant in all the streaks it formed. In reality, asteroids rotate and often have nonspherical shapes, which causes the apparent brightness to vary. Most of the asteroids larger than approximately 150 m have rotation periods longer than two hours, but smaller asteroids can rotate much faster (Pravec et al. 2002). Therefore, as the observation time for a pointing is 70 min, larger asteroids can show variation in brightness mostly between different dithers. Smaller asteroids can show significant variation even within a single dither, resulting in a streak with periodic brightness variation, and if the brightness is close to the limiting magnitude, a





**Fig. 4.** Errors in angles, lengths, and coordinates of *StreakDet* finds as a function of magnitude and length. The y-axes of the two uppermost and the two lowermost plots are in logarithmic scale. The y-axes of the middle plots are in symmetric logarithmic scale, where the values between  $-10^{-1}$  and  $10^{-1}$  are shown in linear scale. In the middle plots, showing length errors, negative y-values mean that the corresponding lengths given by *StreakDet* are shorter than ground-truth lengths, whereas positive values mean that lengths given by *StreakDet* are longer than ground truth. The 70th percentile length errors in the middle plots are technically 30th percentile errors, since the length errors are typically negative.

dashed line. *StreakDet* can link a dashed line into a coherent streak feature, but the variation in apparent magnitude can make the detection more difficult. Measuring the significance of this effect on the detection results requires more realistic simulations and remains for future work.

## 5.2. Errors

Figure 4 shows the angle, length, and coordinate errors of *StreakDet* detections as functions of the true magnitude and true streak length. The fitted angles of the streaks are typically accurate when compared to the ground truth. For the brightest streaks, in the magnitude range of 20–21, the average angle error is around 0.06 degrees, and the median angle error is 0.03 degrees. For fainter streaks, the angle errors increase

steadily due to a less accurate streak fitting, and after magnitude 24, the average error is above 1 degree. For all magnitudes, the average angle error is 0.22 degrees, and the median angle error is 0.08 degrees. After magnitude 25, there were just a few individual detections, which causes the median and mean errors to be identical.

For shorter streaks, the angle error is larger than for long streaks due to a better streak fitting for long streaks. In the error plots that are plotted as a function of sky motion, there appears to be a plateau at the very shortest length range, but this is again due to the fact that there are very few or zero found streaks at those points.

The fitted lengths of the *StreakDet* finds are also quite accurate for bright streaks. For streaks brighter than 21 magnitudes, the average error in streak length is  $-0.6\%$ , that

is, the streaks found by StreakDet are slightly (at sub-pixel level) shorter than the ground truth. The median error for magnitudes smaller than 21 is 0.04%, so that the lengths of the streaks found are very slightly longer than the ground-truth streaks. Again, the error and its variance start to increase for fainter streaks due to less accurate streak fitting. The sign of the length error as a function of magnitude changes between magnitudes 22.6 and 24, and again around magnitude 25. When going to fainter streaks, the variance of the errors increase due to a less accurate streak fitting, so the change of the error sign is likely just an effect of randomness. This is especially true after magnitudes 24 and even more so after 25, where the number of found streaks is very small and the streaks are very faint. The very small number of detections for the faintest streaks causes the median and mean errors to become identical in the plot.

For long streaks, StreakDet sometimes finds two shorter line segments of the ground-truth streak instead of the entire streak. In addition, some streaks fall partially outside of the image, and only a part of them are visible on the CCD. This causes the found length to be shorter than the ground-truth length. Longer streaks are more likely to continue outside the image. These effects combined are responsible for the median and mean length errors turning negative for streaks longer than approximately 60 pixels ( $40 \text{ arcsec h}^{-1}$ ). For long streaks, the median length error is negative at sub-pixel level, whereas the mean length error approaches  $-10\%$ . For short streaks, the median and mean length errors are positive at sub-pixel level.

In other words, both angle and length errors increase for faint streaks. For short streaks, angle errors are larger, but relative length errors are smaller. For the long streaks, the opposite is true, so that angle errors are small, but relative length errors are larger. A notable detail is that while the mean error can be large, the corresponding median error is typically much smaller. This means that most streak parameters defined by StreakDet are quite close to the ground truth, and for a small number of streaks, the parameters are off by a large margin. In addition to the causes for larger errors discussed above, a general case for a sizable error is that some streaks fall close to or on top of galaxies or other objects, which can interfere with the streak fitting.

The coordinate error is defined as the difference between the ground-truth coordinates for the middle point of the streak and the corresponding coordinates given by StreakDet. For the brightest streaks, between magnitudes 20 and 21, the average coordinate error is 0.58 pixels, and the median coordinate error is 0.10 pixels, corresponding to 58 and 10 milliarcseconds, respectively. The coordinate error increases for both faint and long streaks. For all magnitudes, the average coordinate error is 1.92 pixels (192 mas), and the median coordinate error is 0.19 pixels (19 mas).

The primary source of the coordinate error is the length error. When StreakDet detects the length of the streak inaccurately, the middle point of the streak is also detected inaccurately. When taking only streaks whose length error is less than 0.1 pixels into account (22% of all detections), the average coordinate error for all magnitudes decreases to 0.15 pixels (15 mas), and the median error decreases to 0.10 pixels (10 mas).

We tested StreakDet without astrometry and photometry at this point, as it was originally programmed to use UCAC4, and *Gaia* data are available for *Euclid*. Therefore, there are no estimates for apparent magnitude errors.

### 5.3. Multi-streak analysis

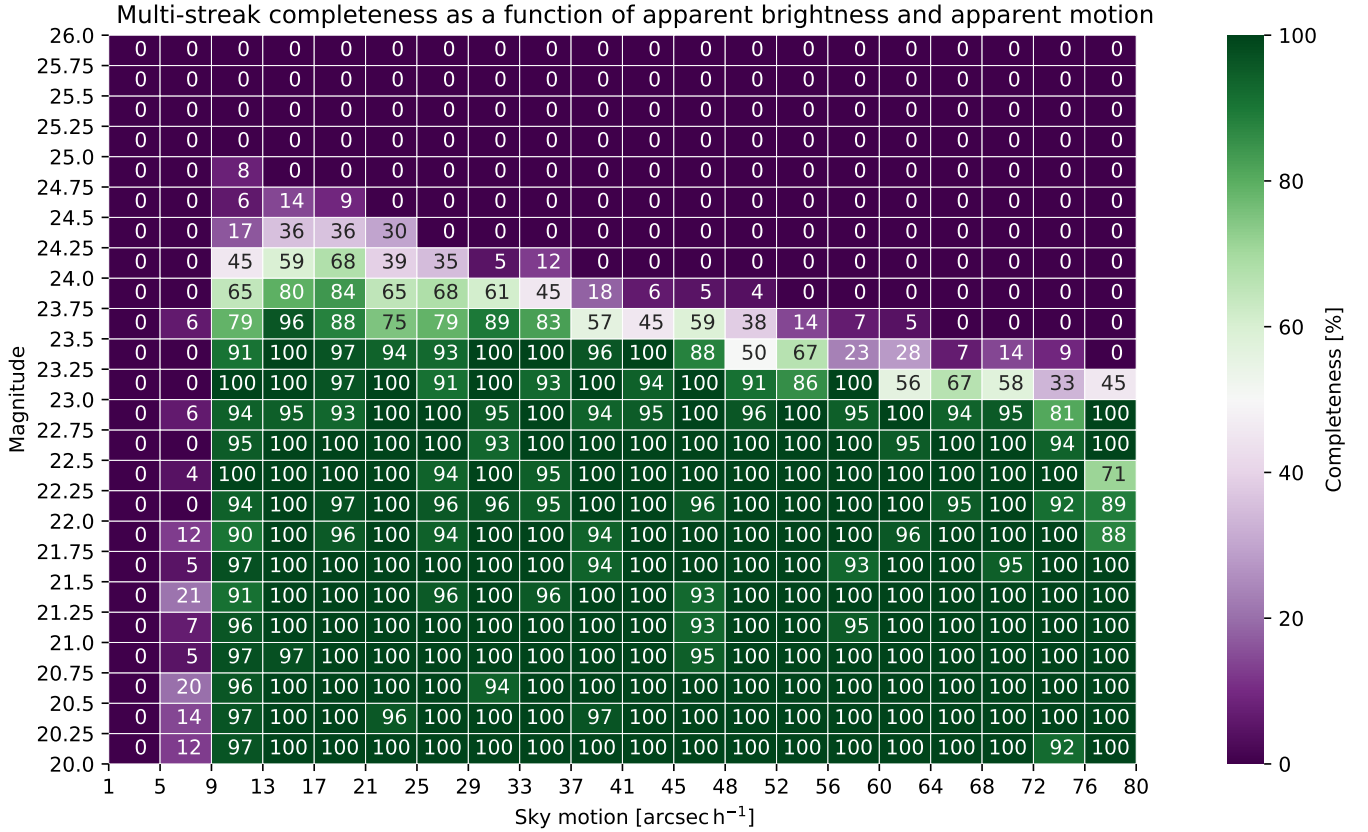
Since the settings were optimized to maximize the detection percentage, when analyzing just the individual streaks, the number of false positives is high, and there are approximately 30 false positives for each true positive. Multi-streak analysis manages to discard most of the false positives, while maintaining most of the true positives. The multi-streak analysis results are shown in Fig. 5 and in Tables 2 and 3. The completeness is defined here as the fraction of multi-streaks found among the ground-truth multi-streaks that have visible streaks in at least two dithers. In other words, ground-truth objects that are visible in only one dither are ignored. These objects correspond to approximately 9% of the simulated objects. If they were included, all the completeness values shown in the heat map would be approximately 9% lower.

The parameters of multi-streak analysis are the following, with values we used in the analysis in the parentheses: minimum number of streaks in a multi-streak (2), maximum length difference between streaks (45%), maximum angle difference between streaks ( $3^\circ$ ), maximum angle difference between the single streaks and the common multi-streak line ( $3^\circ$ ), maximum flux difference between streaks (29%), and maximum difference in PSF sigma values (44%). This set of parameters was found to retain virtually all of the true positives, while decreasing the number of false positives by three orders of magnitude to approximately 4% of all StreakDet finds. The false-positive multi-streaks are typically caused by two individual false-positive streaks that happen to be approximately along the same line in two dithers and otherwise fulfill the filtering parameters.

The number of false positives can be decreased further by using stricter parameters, but the number of true positives starts to decrease faster than the number of false positives below a 4% false-positive level. The most straightforward way to achieve virtually zero false positives is to require at least three linked streaks instead of two. This appears to guarantee 100% purity, but the downside is a fairly large drop in the number of true positives, from 55.66% to 44.36% completeness in a full test set containing all tested streak lengths and magnitudes. The multi-streak analysis is a separate post-processing step, so it can be run multiple times with different settings, without the need to run StreakDet again. This way, multiple SSO data sets can be offered, for example, one with zero false positives but smaller completeness, and another with a small number of false positives but higher completeness. Currently, the multi-streak analysis takes approximately one minute to run for one pointing, 144 CCDs. With more optimization, the running time can be further reduced, if needed.

When comparing the top right and bottom right images in Fig. 1, it seems that a substantial portion of the false positives are results of the streaks caused by the CTI shadows of the removed cosmic rays. As the OU-VIS data reduction pipeline considerably reduces the CTI effects in the images, the number of false-positive streaks should also decrease, increasing purity.

The fifth VIS exposure, which was not simulated in our data, can also help increase purity and completeness. However, the detectability of SSOs in the additional VIS exposure with StreakDet is limited to objects moving faster than  $\sim 60 \text{ arcsec h}^{-1}$  because the exposure time of the additional VIS frame is only 116 s, instead of 565 s, resulting in much shorter streaks.



**Fig. 5.** Multi-streak detection percentage as functions of apparent magnitude and apparent motion. The values on both axes show the limits of the bins. On average, there are 21 ground-truth multi-streaks per bin. The completeness shown is defined as the number of true-positive linked multi-streaks divided by the number of ground-truth multi-streaks consisting of at least two linked streaks. In other words, ground-truth streaks appearing in only one dither are ignored. If they were included, all the completeness values shown would be approximately 9% lower.

**Table 2.** Multi-streak analysis results, combined for all tested data sets. Semi-hits refer to linked multi-streaks that contain both true and false-positive streaks.

Type	Number	Percentage
Ground-truth multi-streaks	9737	
StreakDet multi-streak finds	5682	
Hits	5420	55.66%
Semi-hits	18	0.18%
Total hits	5438	55.85%
Duplicates	17	0.17%
Misses	4299	44.15%
False positives	227	4.00% of SD finds

**Notes.** The table includes all the data underlying Fig. 5, including all the zero-bins, which explains the total completeness of only 55.66%.

## 6. Conclusions and future work

In the simulated *Euclid* data, StreakDet finds a streak in at least one dither for 99.1% of SSOs brighter than 23 magnitudes, with streak lengths longer than 15 pixels ( $10 \text{ arcsec h}^{-1}$ ), and it finds 96.9% of all the individual streaks caused by these objects. For streaks fainter than magnitude 23, the finding percentage starts to decrease and reaches zero at magnitude 25. This is to be expected, given the limiting magnitudes of VIS. Still,

**Table 3.** Detailed detection percentages of multi-streak analysis. GT refers to ground truth and SD refers to StreakDet.

<i>GT multi-streaks with 4 parts</i>	4009	41.17% of GTs
Of which StreakDet found	2285	57.00%
With 4 SD hits	1612	40.21%
With 3 SD hits	491	12.25%
With 2 SD hits	182	4.54%
<i>GT multi-streaks with 3 parts</i>	4828	49.58% of GTs
Of which StreakDet found	2692	55.76%
With 3 SD hits	2211	45.80%
With 2 SD hits	481	9.96%
<i>GT multi-streaks with 2 parts</i>	900	9.24% of GTs
Of which StreakDet found	443	49.22%
With 2 SD hits	443	49.22%

there appears to be some room for improvement in the detection percentage above magnitude 23 since StreakDet fails to detect some faint streaks that are still visible to the naked eye, and it finds some streaks of similar brightness, but not others. For streaks shorter than 13 pixels ( $8 \text{ arcsec h}^{-1}$ ), the detection percentage declines rapidly, approaching zero at streak lengths below 9 pixels ( $6 \text{ arcsec h}^{-1}$ ). The multi-streak analysis run on the final StreakDet results worked well in retaining most of the true positives, while removing most or, with stricter settings, all of the false positives.

Going forward, the areas with the greatest potential for improving the overall detection ability of asteroids in *Euclid* data are improving the segmentation phase of StreakDet, making it more readily able to detect short and/or faint streaks. Improving the detection percentage for faint streaks is especially important, because most of the asteroids appearing in *Euclid* images are close to the limiting magnitude. The simulated data used in this work contained too few faint galaxies, compared to the real sky. This could somewhat change the expected StreakDet detection percentage for streaks above magnitude 24. However, as StreakDet currently detects very few of the streaks that are that faint, the effect on results should be small, and improving the detection of faint streaks with realistic faint galaxy density remains for future work in any case. If possible, it would also be useful to optimize the StreakDet code so that it runs more rapidly, because currently, it takes approximately one minute to analyze one  $4\text{k} \times 4\text{k}$  image. Another option is to develop an advanced deep learning model, such as a convolutional neural network, that is capable of directly returning the coordinates of the asteroids in the images. One possible option is to combine these two approaches. From a practical point of view, it is probably necessary to collect at least some non-simulated *Euclid* training data for a machine-learning approach using a method such as StreakDet. Another prospect is studying the analysis of asteroids in the NISP images, in terms of detection and measurement of the spectral energy distributions.

**Acknowledgements.** We thank the anonymous referee for constructive comments. M.P. and M.G. acknowledge funding from the Academy of Finland (projects #316292 and #299543). The Euclid Consortium acknowledges the European Space Agency and the support of a number of agencies and institutes that have supported the development of *Euclid*. A detailed complete list is available on the *Euclid* web site (<http://www.euclid-ec.org>). In particular, the Academy of Finland, the Agenzia Spaziale Italiana, the Belgian Science Policy, the Canadian Euclid Consortium, the Centre National d'Etudes Spatiales, the Deutsches Zentrum für Luft- und Raumfahrt, the Danish Space Research Institute, the Fundação para a Ciência e a Tecnologia, the Ministerio de Economía y Competitividad, the National Aeronautics and Space Administration, the Nederlandse Onderzoekschool Voor Astronomie, the Norwegian Space Agency, the Romanian Space Agency, the State Secretariat for Education, Research and Innovation (SERI) at the Swiss Space Office (SSO), and the United Kingdom Space Agency.

## References

- Amendola, L., Appleby, S., Avgoustidis, A., et al. 2018, *Liv. Rev. Rel.*, **21**, 2
- Carry, B. 2018, *A&A*, **609**, A113
- Cook, N. V., Ragozzine, D., Granvik, M., & Stephens, D. C. 2016, *ApJ*, **825**, 51
- Crocce, M., Castander, F., Gaztanaga, E., Fosalba, P., & Carretero, J. 2015, *MNRAS*, **453**, 1513
- Cropper, M., Pottinger, S., Azzollini, R., et al. 2018, *Space Telescopes and Instrumentation 2018: Optical, Infrared, and Millimeter Wave* (SPIE), 10698, 1069828
- Dubath, P., Apostolakis, N., Bonchi, A., et al. 2016, *Proc. Int. Astron. Union*, **12**, 73
- Duev, D. A., Mahabal, A., Ye, Q., et al. 2019, *MNRAS*, **486**, 4158
- Gaffey, M., Cloutis, E., Kelley, M., & Reed, K. 2002, *Asteroids III* (Tucson: The University of Arizona Press), 183
- Gaia Collaboration (Prusti, T., et al.) 2016, *A&A*, **595**, A1
- Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, ArXiv eprint [arXiv:1110.3193]
- Lieu, M., Conversi, L., Altieri, B., & Carry, B. 2019, *MNRAS*, **485**, 5831
- Maciaszek, T., Ealet, A., Jahnke, K., et al. 2016, *Space Telescopes and Instrumentation 2016: Optical, Infrared, and Millimeter Wave* (SPIE), 9904, 99040T
- Mahlke, M., Bouy, H., Altieri, B., et al. 2018, *A&A*, **610**, A21
- McCully, C., Crawford, S., Kovacs, G., et al. 2018, <https://doi.org/10.5281/zenodo.1482019>
- Muñinonen, K., Virtanen, J., Granvik, M., & Laakso, T. 2006, *MNRAS*, **368**, 809
- Nir, G., Zackay, B., & Ofek, E. O. 2018, *AJ*, **156**, 229
- Pravec, P., Harris, A. W., & Michalowski, T. 2002, *Asteroids III* (Tucson: The University of Arizona Press), 113
- Racca, G. D., Laureijs, R., Stagnaro, L., et al. 2016, *Space Telescopes and Instrumentation 2016: Optical, Infrared, and Millimeter Wave* (SPIE), 9904, 990400
- Reddy, V., Dunn, T. L., Thomas, C. A., Moskovitz, N. A., & Burbine, T. H. 2015, in *Asteroids IV*, eds. P. Michel, F. E. DeMeo, & W. F. Bottke, 43
- Shrivastava, A., Pfister, T., Tuzel, O., et al. 2017, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2107
- Snodgrass, C., Carry, B., Berthier, J., et al. 2018, ArXiv e-prints [arXiv:1812.00607]
- Van Dokkum, P. G. 2001, *PASP*, **113**, 1420
- Venancio, L. M. G., Laureijs, R., Lorenz, J., et al. 2014, *Space Telescopes and Instrumentation 2014: Optical, Infrared, and Millimeter Wave* (SPIE), 9143, 91430I
- Virtanen, J., Poikonen, J., Sääntti, T., et al. 2016, *Adv. Space Res.*, **57**, 1607
- Waszczak, A., Prince, T. A., Laher, R., et al. 2017, *PASP*, **129**, 034402
- <sup>1</sup> Department of Physics, PO Box 64, 00014 University of Helsinki, Finland  
e-mail: mikko.pontinen@helsinki.fi
- <sup>2</sup> Asteroid Engineering Laboratory, Onboard Space Systems, Luleå University of Technology, Box 848, 98128 Kiruna, Sweden
- <sup>3</sup> INFN, Sezione di Lecce, Via per Arnesano, CP-193, 73100 Lecce, Italy
- <sup>4</sup> Department of Mathematics and Physics E. De Giorgi, University of Salento, Via per Arnesano, CP-193, 73100 Lecce, Italy
- <sup>5</sup> European Space Agency/ESRIN, Largo Galileo Galilei 1, 00044 Frascati, Roma, Italy
- <sup>6</sup> ESAC/ESA, Camino Bajo del Castillo, s/n, Urb. Villafranca del Castillo, 28692 Villanueva de la Cañada, Madrid, Spain
- <sup>7</sup> INAF-Osservatorio di Astrofisica e Scienza dello Spazio di Bologna, Via Piero Gobetti 93/3, 40129 Bologna, Italy
- <sup>8</sup> Max Planck Institute for Extraterrestrial Physics, Giessenbachstr. 1, 85748 Garching, Germany
- <sup>9</sup> INAF-Osservatorio Astrofisico di Torino, Via Osservatorio 20, 10025 Pino Torinese (TO), Italy
- <sup>10</sup> INAF-Osservatorio Astronomico di Capodimonte, Via Moirariello 16, 80131 Napoli, Italy
- <sup>11</sup> Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, 08193 Bellaterra (Barcelona), Spain
- <sup>12</sup> Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, Laboratoire Lagrange, Bd de l'Observatoire, CS 34229, 06304 Nice Cedex 4, France
- <sup>13</sup> INAF-Osservatorio Astronomico di Roma, Via Frascati 33, 00078 Monteporzio Catone, Italy
- <sup>14</sup> Centre National d'Etudes Spatiales, Toulouse, France
- <sup>15</sup> Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK
- <sup>16</sup> Mullard Space Science Laboratory, University College London, Holmbury St Mary, Dorking, Surrey RH5 6NT, UK
- <sup>17</sup> INFN-Padova, Via Marzolo 8, 35131 Padova, Italy
- <sup>18</sup> INAF-Osservatorio Astronomico di Trieste, Via G. B. Tiepolo 11, 34131 Trieste, Italy
- <sup>19</sup> INAF-IASF Milano, Via Alfonso Corti 12, 20133 Milano, Italy
- <sup>20</sup> von Hoerner & Sulger GmbH, Schloßplatz 8, 68723 Schwetzingen, Germany
- <sup>21</sup> Universitäts-Sternwarte München, Fakultät für Physik, Ludwig-Maximilians-Universität München, Scheinerstrasse 1, 81679 München, Germany
- <sup>22</sup> Max-Planck-Institut für Astronomie, Königstuhl 17, 69117 Heidelberg, Germany
- <sup>23</sup> Aix-Marseille Univ., CNRS/IN2P3, CPPM, Marseille, France
- <sup>24</sup> Univ. Lyon, Univ. Claude Bernard Lyon 1, CNRS/IN2P3, IP2I Lyon, UMR 5822, 69622 Villeurbanne, France



- <sup>25</sup> Université de Genève, Département de Physique Théorique and Centre for Astroparticle Physics, 24 quai Ernest-Ansermet, 1211 Genève 4, Switzerland
- <sup>26</sup> European Space Agency/ESTEC, Keplerlaan 1, 2201 AZ Noordwijk, The Netherlands
- <sup>27</sup> Institute of Theoretical Astrophysics, University of Oslo, PO Box 1029 Blindern, 0315 Oslo, Norway
- <sup>28</sup> Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, 08193 Barcelona, Spain
- <sup>29</sup> Institut d'Estudis Espacials de Catalunya (IEEC), 08034 Barcelona, Spain
- <sup>30</sup> Argelander-Institut für Astronomie, Universität Bonn, Auf dem Hügel 71, 53121 Bonn, Germany
- <sup>31</sup> Centre for Extragalactic Astronomy, Department of Physics, Durham University, South Road, Durham DH1 3LE, UK
- <sup>32</sup> Observatoire de Sauverny, Ecole Polytechnique Fédérale de Lausanne, 1290 Versoix, Switzerland
- <sup>33</sup> Dipartimento di Fisica e Astronomia, Università di Bologna, Via Gobetti 93/2, 40129 Bologna, Italy
- <sup>34</sup> INFN-Bologna, Via Irnerio 46, 40126 Bologna, Italy
- <sup>35</sup> Department of Astronomy, University of Geneva, Ch. d'Écogia 16, 1290 Versoix, Switzerland
- <sup>36</sup> AIM, CEA, CNRS, Université Paris-Saclay, Université Paris Diderot, Sorbonne Paris Cité, 91191 Gif-sur-Yvette, France
- <sup>37</sup> Space Science Data Center, Italian Space Agency, via del Politecnico snc, 00133 Roma, Italy
- <sup>38</sup> INFN-Sezione di Bologna, Viale Berti Pichat 6/2, 40127 Bologna, Italy
- <sup>39</sup> Instituto de Astrofísica e Ciências do Espaço, Faculdade de Ciências, Universidade de Lisboa, Tapada da Ajuda, 1349-018 Lisboa, Portugal
- <sup>40</sup> Departamento de Física, Faculdade de Ciências, Universidade de Lisboa, Edifício C8, Campo Grande, 1749-016 Lisboa, Portugal
- <sup>41</sup> Universidad Politécnica de Cartagena, Departamento de Electrónica y Tecnología de Computadoras, 30202 Cartagena, Spain
- <sup>42</sup> Infrared Processing and Analysis Center, California Institute of Technology, Pasadena, CA 91125, USA